



Universitat
Autònoma
de Barcelona



APLICACIÓ WEB DE NOTÍCIES

Memòria del Projecte Fi de Carrera
d'Enginyeria en Informàtica
realitzat per

Lluís Llopart i Vidal

i dirigit per

Anna Barbara Morajko

Bellaterra, 12 de juny de 2008

Català

Aquesta memòria tracta sobre el procediment de creació d'una aplicació web de notícies. Està dividida en 3 zones, una on usuaris amb permisos d'administració poden penjar notícies per ser visualitzades per tothom, una altra que s'hi accedeix si s'és usuari registrat i permet visualitzar notícies d'altres servidors mitjançant el format de dades RSS, i un tercer apartat de gestió administrativa, incorporar noves notícies, modificar-ne de presents o introduir noves pàgines web que continguin notícies. Els usuaris registrats podran seleccionar el diaris dels quals rebran informació, així com especificar quines temàtiques prefereixen en la cerca de notícies.

Español

Esta memoria trata sobre el procedimiento de creación de una aplicación web de noticias. Está dividida en 3 zonas, una donde usuarios con permisos de administración pueden colgar noticias para ser visualizadas por todos, otra que se accede si se es usuario registrado y que permite visualizar noticias de otros servidores mediante el formato de datos RSS, y un tercer apartado de gestión administrativa, incorporar nuevas noticias, modificar las presentes o introducir nuevas páginas web que contengan noticias. Los usuarios registrados podrán seleccionar los periódicos de los cuales recibirán información, así como especificar qué temáticas prefieren en la búsqueda de noticias.

English

This report is about the creation procedure of a news web application. It is divided into 3 areas. In the first one, users with administration permit are allowed to hang news to be visualized by everybody; the second one, which is accessed only by a registered user and allows visualizing news from other servers through the data format RSS, and, finally, a third area of administrative management in order to introduce and modify news or to introduce further web pages containing news. Registered users will be allowed to select the newspapers from which they will receive information as well as to specify which news search topic they prefer.

El sotasignat, **Anna Barbara Morajko**

Professor/a de l'Escola Tècnica Superior d'Enginyeria de la UAB,

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per en **Lluís Llopert i Vidal**

I per tal que consti firma la present.

Signat:

Bellaterra, 12 de juny de 2008

ÍNDIX

1. INTRODUCCIÓ	3
1.1 Notícies	3
1.2 Generadors.....	4
1.3 Agregadors	5
1.4 Objectius	6
1.5 Organització de la memòria	7
2. REQUERIMENTS.....	8
2.1 Públic al qual va dirigit.....	8
2.1.1 Visitant	8
2.1.2 Usuari Registrat	8
2.1.3 Administrador	8
2.2 Mètodes de recepció de les notícies.....	9
2.2.1 Notícies pròpies	9
2.2.2 Notícies externes	9
2.3 Funcions que ha de fer l'aplicació.....	9
2.3.1 Visualització d'una notícia	9
2.3.2 Visualització de les últimes cinc notícies	11
2.3.3 Menú superior amb els temes predefinits	12
2.3.4 Cerca per paraula	13
2.3.5 Introducció d'una notícia	14
2.3.6 Modificar una notícia.....	15
2.4 Esquema general.....	15
2.4.1 Frame estàtic superior.....	16
2.4.2 Frame dinàmic central	16
2.4.3 Frame dinàmic inferior.....	17
3. ARQUITECTURA	18
3.1 Plataforma WAMP.....	18
3.1.1 Windows (Sistema Operatiu)	18

3.1.2 Apache (Servidor Web).....	19
3.1.3 MySQL (Base de Dades).....	21
3.1.4 PHP (Llenguatge de Programació)	22
3.2 Software utilitzat	24
3.2.1 phpMyAdmin.....	25
3.2.2 Macromedia Dreamweaver.....	25
3.2.3 Corel Photo Paint.....	25
3.2.4 Internet Explorer i Mozilla Firefox	26
3.3 Format de dades RSS	26
3.4 Interacció global.....	28
4. DISSENY I IMPLEMENTACIÓ.....	30
4.1 Diagrama E-R	30
4.1.1 Part d'usuari visitant.....	31
4.1.2 Part d'usuari registrat.....	31
4.1.3 Part d'usuari administrador.....	31
4.2 Disseny funcional.....	32
4.2.1 Introducció login i password.....	32
4.2.2 Visualització de notícies pròpies.....	33
4.2.3 Visualització de notícies externes	34
4.2.4 Modificació de la selecció	37
4.2.5 Cerca de paraules	42
4.2.6 Cerca per opcions avançades	43
4.2.7 Inserir nous diaris a l'aplicació	44
4.2.8 Guardar notícies a l'aplicació	45
5. CONCLUSIONS.....	46
5.1 Treball futur	48
BIBLIOGRAFIA	49
Bibliografia general.....	49
Bibliografia específica	49

1. INTRODUCCIÓ

El títol d'aquest projecte, Aplicació web de notícies, és una definició molt exacte del que serà al final. Més endavant, a l'apartat 4 d'aquest capítol, s'analitzaran els objectius en sí, però per fer una breu síntesi a la memòria s'ha de dir que es tracta de poder recollir notícies de diferents fonts d'informació i visualitzar-les segons el que l'usuari desitgi integrades dins d'una aplicació uniforme.

1.1 Notícies

La definició de notícia al Gran Diccionari de l'Enciclopèdia Catalana [1] és:

1. Coneixença que hom té d'alguna cosa.
2. Escrit destinat a donar coneixença d'un punt d'història, de ciència, etc...
3. Anunci d'alguna cosa esdevinguda de poc fet a qui no en té coneixement.
4. Informació sobre l'estat, la situació, etc, d'una persona o d'una cosa de la qual fa algun temps que hom no sap res.

Per tant després d'escriure la definició de diccionari de notícia es pot extreure que les notícies han sigut durant segles i segles el mitjà de transmissió de la informació. Les notícies s'han propagat de diferent manera, primerament de forma manuscrita, posteriorment i gràcies a l'impremta amb edicions i tirades més nombroses. Les editorials de diaris en són un exemple ben clar, però l'abast on hi pot incidir és relativament petit, sobretot si la central on es disposen les impressores no està distribuït pel territori. Amb el temps les notícies s'han difós gràcies als mitjans audiovisuals, ja sigui en un primer moment la ràdio i posteriorment la televisió, la qual cosa feia que les notícies es distribuïssin a més velocitat pel territori, però ha sigut internet el que ha revolucionat el món de la informació. Accelerant el procés de globalització general, i fent que les notícies es puguin llegir de forma quasi instantània des de qualsevol lloc del planeta.

1.2 Generadors de notícies

Per una banda internet està ple de pàgines web de notícies, tots els diaris digitals, totes les agències de notícies, els portals de mitjans de comunicació tenen la seva plana web amb notícies actualitzades cada cop amb una freqüència més gran. I la gran majoria d'elles tenen activat el sistema RSS [2], explicat en detall al capítol 3 on es tracta l'Arquitectura de l'aplicació, per poder compartir la informació entre ells amb una gran facilitat i rapidesa. De totes maneres aquests portals gairebé sempre tenen informació que prové dels seus empleats, de les seves redaccions i poques vegades ofereixen informació que algú altra ha recollit. Jo em basaré amb els portals dels diaris digitals. Per exemple La Vanguardia té la seva adreça "normal" amb tota la informació que la gent pot veure a <http://www.lavanguardia.es/> però a més a més té una altra pàgina, que és la següent: <http://www.lavanguardia.es/rss/index.html> on apareix en forma d'índex una sèrie de temàtiques per poder accedir a tota la informació reduïda com a titulars. Aquesta dualitat de pàgines la tenen la majoria de diaris digitals i s'accedeix a la segona pitjant sobre la icona RSS, mostrada tot seguit a la Fig. 1 i que normalment es troba a la part dreta de la barra d'adreces del navegador o a la part inferior de la pàgina principal.



Fig. 1 – Logotip RSS

A part de les pàgines d'informació, podríem dir-n'hi oficial, de les grans empreses, de les editorials de diaris o de les agències de notícies, existeixen una sèrie de pàgines generadores de notícies fetes per usuaris anònims. Aquestes pàgines estan en creixement constant degut a la facilitat en l'actualització i velocitat en la difusió, s'anomenen blogs. Pàgines personals, on cada usuari va posant les notícies que creu oportú en el moment que vulgui. Amb enllaços RSS s'aconsegueix distribuir la informació a una velocitat molt alta ja que automàticament el text córrer per la xarxa.

1.3 Agregadors de notícies

Un agregador és un tipus de software per a subscriure's a canals de notícies en formats RSS, Atom i altres derivats de XML/RDF. L'agregador reuneix les notícies o històries publicades en els llocs web sindicats escollits, i mostra les novetats o modificacions que s'han produït en aquests canals, és a dir, avisa de quines webs han incorporat contingut nou des de l'última lectura que s'hi hagi fet i quin és aquest contingut. Aquesta informació és la que es coneix amb el nom de feed. Utilitzant el sistema de la tecnologia RSS hi ha una sèrie de programes que s'anomenen Agregadors, els quals permeten rebre els titulars de les notícies de les pàgines web que posis al programa. Tots ells es poden classificar en tres categories.

- Agregadors d'escriptori: S'instal·len a l'ordinador de l'usuari. Mitjançant un programa s'accedeix a la informació. Hi ha moltes empreses i programes per la xarxa, però un dels agregadors d'escriptori més habituals és el *FeedReader*, encara que també hi ha altres agregadors com *RSSOwl*, *RSSReader* o *AmphetaDesk*, entre d'altres.
- Agregadors en línia: No necessiten instal·lació per part de l'usuari. Normalment n'hi ha prou donant-se d'alta al lloc web per poder-lo utilitzar. Alguns exemples són: *Bloglines*, *Feedness*, *Newsgator*, *Google Reader* o *Mi Yahoo!*
- Agregadors plug-ins: Alguns navegadors i gestors de correu els inclouen als seus programes com a valor afegit. *Mozilla Firefox* i *Thunderbird* en són uns exemples.

Algunes empreses com ara *Feedster*, *Plazoo*, *Retronimo*, *YourFeeds*, *Amazon*, *AllConsuming*, *NewsIsFree*, *sindic8*, *Blogdigger*, i un llarg etcètera estan explorant noves formes d'ús i distribució de la informació.

Tots aquests programes accepten una adreça d’RSS i mostren els titulars de les notícies que conté el servidor que se li ha passat. Si interessa veure les notícies senceres hi ha un enllaç a la notícia, la qual es troba al servidor original.

Tots els documents XML que reben els agregadors consten com a mínim de tres camps fixes: un títol, una descripció de la notícia i un enllaç a la notícia sencera que està allotjada a cada servidor d’origen.

1.4 Objectius

L’objectiu d’aquest projecte és poder tenir, tal com diu el títol, una aplicació web de notícies, això significa tenir un portal web que permeti rebre notícies segons les característiques que l’usuari desitgi, determinar de quins llocs es vol rebre, o quin tipus de notícies es vol tenir.

Els objectius principals d’aquest projecte són:

- Tenir un portal de notícies adaptat a les necessitats de tothom qui hi entri.
- Distingir 3 zones al portal, una d’usuaris visitants, una d’usuaris registrats i una tercera d’administració.
- La zona d’usuaris visitants només ha de permetre visualitzar les notícies que el nostre portal contingui a la nostra base de dades.
- La zona d’usuaris registrats ha de permetre importar notícies de diverses pàgines web mitjançant la tecnologia RSS.
- Seleccionar les pàgines web d’on es volen treure les notícies.
- Seleccionar temàtiques específiques per rebre només unes certes notícies.
- La zona d’usuaris amb perfil d’administradors ha de permetre introduir notícies segons la necessitat que es tingui.

1.5 Organització de la memòria

En aquest document s'intenta explicar el perquè s'ha dut a terme aquest projecte, quines són les aplicacions actuals, així com tot allò que es fa, com es fa i amb què es fa, seguidament es detalla l'estructura de la memòria.

- En el segon capítol s'enumera a qui va dirigida l'aplicació i es descriuen totes les funcions que ha de fer aquesta. Es fa un èmfasi especial en els requeriments que calen per poder dur a terme aquest projecte.
- Al capítol tercer s'explica amb quina arquitectura es treballa, quin software s'ha utilitzat per poder dur a terme aquest projecte.
- Al quart capítol s'explica com s'ha dut a terme l'aplicació, les decisions preses que n'afecten al disseny i implementació.
- Al capítol cinquè finalment es mostra una valoració final personal del projecte i a descriure'n les millores que se'n podrien fer en un futur.
- El darrer capítol de la memòria és la bibliografia que s'ha usat durant aquest període.

2. REQUERIMIENTS

2.1 Públic al qual va dirigit

2.1.1 Visitant

Aquest usuari tan sols podrà visualitzar les notícies que tinguem penjades en el servidor, les notícies pròpies, detallades en el següent apartat de la memòria. No ha d'introduir cap contrasenya però té l'accés restringit, no pot accedir a la informació procedent d'altres webs. Un visitant es pot registrar directament des del portal si així ho desitja, i passa a convertir-se en usuari registrat amb els drets que es descriuen a continuació.

2.1.2 Usuari Registrat

L'usuari registrat accedirà al portal mitjançant un login i un password. Aquest usuari podrà visualitzar les notícies pròpies, i a més podrà configurar al seu gust les característiques que més li convinguin per visualitzar la informació més adequada a les seves necessitats. Aquest usuari tindrà accés a les notícies externes a través de la tecnologia RSS mitjançant el qual rebrà les notícies d'altres portals d'internet. Aquesta tecnologia la detallarem més endavant, però es basa en l'actualització constant d'informació procedent de diferents llocs web.

2.1.3 Administrador

L'administrador és la persona que accedeix al portal mitjançant un login i un password en la part d'administració de la web, i es distingeix dels altres dos perquè aquest pot introduir notícies noves. Les notícies que introdueix l'administrador són les que hem anomenat *notícies pròpies*. Per una altra part també pot afegir diaris a la part de *notícies externes*. És un accés restringit només per al personal que treballi directament relacionat amb la pàgina.

2.2 Mètodes de recepció de les notícies

2.2.1 Notícies pròpies

Per un costat hi ha les notícies que hem anomenat *notícies pròpies*, són aquelles que un administrador ha introduït a la nostra pàgina. Aquestes seran les notícies que apareguin per defecte, quan s'entra al portal, i les notícies que hi haurà ordenades per temàtiques. Aquestes seran les úniques notícies que podran visualitzar els usuaris visitant de l'aplicació. Uns periodistes que treballen per l'aplicació redactaran notícies per exemple que succeeixen a la població, a la comarca o en un cercle relativament reduït.

2.2.2 Notícies externes

L'altre mètode de visualització de les notícies és el denominat RSS, aquest és un sistema de notícies que està implementat en moltes pàgines web i el que intentarem fer és accedir a elles i mostrar-les en la nostra aplicació depenent de les característiques que l'usuari hagi seleccionat. Hi haurà dos tipus de cerques. Es podrà escollir la temàtica de les notícies i/o la procedència d'aquestes. L'usuari registrat tindrà guardada una sèrie d'informació, el tema o temes favorits de les notícies (esports, cultura, economia, societat o política), l'idioma en que es vol rebre la informació, les pàgines d'on es vol rebre la informació i característiques relacionades amb el format del text (color del fons de pantalla, color del text i mida del tipus de lletra). Amb aquest funcionament es mostraran notícies que s'actualitzaran de forma automàtica, quan el servidor que les emet les afegeixi a la seva pàgina. Un exemple d'aquestes notícies pot ser una crònica d'esports del diari El Periódico.

2.3 Funcions que ha de fer el nostre portal

2.3.1 Visualització d'una notícia

Per a visualitzar una notícia hi haurà diferents camps amb la informació respectiva. Anem a descriure'n els seus tipus de dades.

Títol - Títol de la notícia, camp de text amb una llargada de 100 caràcters.

Subtítol - Subtítol, camp de text amb una llargada de 200 caràcters.

Cos - Cos de la notícia, camp amb text llarg.

Data - Data en que la notícia s'ha produït, de text, 10 caràcters.

Hora - Hora en que ha sigut penjada, text de 5 caràcters.

Autor de la notícia si es tracta d'una notícia pròpia o lloc de procedència si la notícia que s'està visualitzant procedeix d'un altre portal mitjançant la tecnologia RSS.

Link - Per a poder enllaçar amb la notícia original si se tracta de RSS.

Evidentment la forma en que es visualitzarà serà amb la mida de la lletra de forma decreixent segons l'ordre d'importància (Títol, subtítol, cos). La data i l'hora s'incorporaran a la part superior de forma més petita i en un color gris. També es mostrarà la temàtica o gènere relacionat amb la notícia, segons si tracta d'una notícia de (esports, política, economia, societat o cultura).

La notícia es visualitzarà dins de la pàgina principal, jugant amb frames, veure Fig. 2. Els frames són les divisions que hi ha a la pàgina, nosaltres crearem 3 particions horitzontals. N'hi haurà uns de fixos, el superior —on se situaran els títols i menús per accedir a les diferents temàtiques— i l'inferior —on es trobarà informació estrictament de la pàgina—. El frame central serà variable depenent de la informació que es vulgui, aquest frame serà el més gran ja que és on es troba la informació en sí.

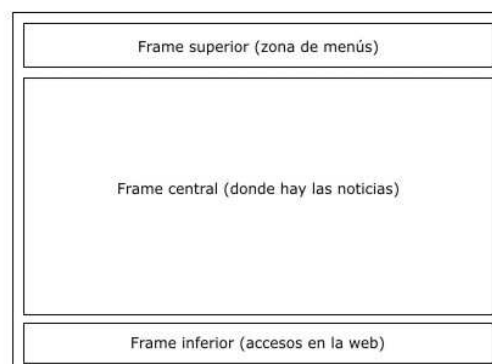


Fig. 2 - Mostra dels 3 frames.

La font de la notícia serà Verdana i en principi el format serà estàtic, no es podrà modificar, tot i que s'introduirà una opció per poder engrandir o empetitir la mida del text del cos de la notícia. A la imatge que apareix a continuació, Fig. 3, es mostra un exemple del que serà la visualització d'una notícia.



Fig. 3 - Visualització d'una notícia.

També es podran introduir imatges a les notícies i si és necessari algun vídeo. La intenció és intercalar les imatges amb el cos de la notícia, col·locant el vídeo, en cas que n'hi hagi una, a la part superior del text, perquè ressalti més i es vegi la seva existència. Evidentment a les notícies externes dependrà del que s'obtingui mitjançant RSS. Les imatges es visualitzaran en una mida relativament reduïda però es podran fer més grans amb una finestra flotant.

2.3.2 Visualització de les últimes cinc notícies

A la pàgina inicial apareixeran les últimes cinc notícies que hi hagi guardades, només es veu el títol i el subtítol a més de la data, hora i tema. Tenint en compte que si s'ha introduït alguna foto relacionada amb la notícia també es visualitzarà a un costat de la pàgina amb una mida més aviat petita, 120 píxels d'alt x 200 píxels d'ample. Si hi ha un vídeo apareixerà amb una icona

indicativa per poder accedir-hi directament i visualitzar-lo en una pàgina flotant.

Aquestes notícies apareixen a mode de breu índex. Des d'aquesta pàgina es pot accedir a la notícia sencera, on ja és possible llegir tota la informació de la notícia i visualitzar totes les fotografies, si n'hi ha més d'una.

2.3.3 Menú superior amb els temes predefinits (Esports / Política / Societat / Economia / Cultura)

Les notícies estan classificades en cinc temàtiques. A la pantalla inicial aquests temes apareixen a la part superior. Es poden seleccionar i així només apareixen aquelles notícies del tema que interessa. Si s'és un usuari registrat i s'han escollit dos o més temes apareixeran les notícies de les temàtiques escollides. Si l'usuari que està demanant la informació és del tipus visitant tan sols podrà seleccionar les notícies tema a tema.

Quan s'entra al portal apareixen els títols en el frame fix superior i les últimes cinc notícies pròpies, que estan al servidor de l'aplicació, en el frame central, així es pot comprovar a la Fig. 4, que apareix a continuació. Si s'accedeix a la pàgina mitjançant un usuari registrat les notícies que apareixeran aniran relacionades amb la informació específica que cada usuari hagi guardat prèviament.

Una vegada seleccionat el tema, apareix una llista de notícies però només amb el títol i el subtítol, a més de la data i l'hora. Per accedir a la notícia completa només cal fer un click a sobre del títol. Per pantalla es mostra un llistat de cinc notícies amb un dispositiu per passar pàgina i anar veient les altres que hi hagi al servidor.



Fig. 4 - Visualització dels 3 frames

2.3.4 Cerca per paraula

L'aplicació també té un apartat per buscar a les notícies la paraula que es vulgui posar. Donada una o diverses paraules busca la notícia que contingui aquests caràcters, ja sigui al títol, al subtítol o al cos de la notícia. Si s'està fent una cerca en una notícia pròpia el funcionament intern serà una consulta select a la base de dades amb la paraula a buscar, i si es tracta d'una cerca a notícies externes es farà una comprovació de la paraula mentre es fa la recerca dels arxius d'origen, més detallat al capítol quart.

També hi ha un altre mecanisme que són les opcions avançades, on es pot especificar més concretament en quin tema es vol buscar la paraula, entre quines dates, marcant uns intervals de temps o en una data concreta depenent del gust de l'usuari, o en quina part del text (títol, subtítol, cos de la notícia). Igualment es poden fer cerques sense especificar cap paraula, simplement seleccionant algun dels camps anteriorment descrits.

Per tal que es pugui visualitzar millor, s'intentarà destacar la paraula marcant-la en un altre color que el del text, si el text és de color negre, per exemple marcar-la en vermell.

2.3.5 Introducció d'una notícia

Per a poder introduir una notícia s'ha de ser administrador de la pàgina. Un cop introduïts el login i password, i un cop verificats els dos, s'introduiran els camps necessaris d'una notícia, havent-hi uns camps obligatoris, uns altres opcionals, i uns darrers automatitzats.

Comencem pels camps obligatoris a omplir:

Títol, subtítol i cos de la notícia. Evidentment són la base de la notícia i no poden quedar en blanc. També s'indicarà el tema de la notícia, aquest s'escollirà d'una llista finita d'opcions predefinides (esports, cultura, societat, economia i política)

Els camps opcionals seran:

Fotografies (una o més) i vídeos.

Automàticament s'ompliran:

Data i hora, s'agafen de l'ordinador, per tant se suposa que cadascú té la data i l'hora de l'ordinador ben posada. També es guardarà qui ha sigut l'autor de la notícia, com que s'ha entrat amb un identificador únic ja el tenim per usar-lo aquí.

En un principi s'aniran guardant totes les notícies que es vagin introduint. Si l'espai que ocupen comptant que hi poden haver fotografies i vídeos és excessivament gran, es marcarà una data límit per anar-les eliminant. Per exemple un any. En aquest cas s'eliminaran de forma automàtica.

Anem a fer un estudi per comprovar si tots els temes s'actualitzen amb la mateixa freqüència, si hi ha algun tema que es carregui d'informació molt més ràpid que els altres. D'aquesta manera es pot configurar el procés d'eliminació de notícies d'una forma particular per cada temàtica. Per posar un exemple, les notícies d'alguns temes s'eliminaran al cap d'un any, però altres poden durar-ne 3 si la quantitat de notícies és molt inferior.

De les notícies externes que s'obtenen mitjançant la tecnologia RSS se'n farà un estudi per comprovar si val la pena copiar la informació a la base de dades per després llegir-la i mostrar-la per pantalla o simplement es capta del portal extern i es mostra sense haver de guardar tota la informació. Depenent de la quantitat de notícies que arribin es valorarà si se segueix un camí o l'altre.

Després de veure que els arxius XML no contenen imatges ni vídeos que s'hagin de copiar a la base de dades i que per tant l'espai que ocupen les notícies no és molt extens s'ha arribat a la conclusió que es guardarà un historial amb totes les notícies de forma diària per tal de poder fer cerques a notícies antigues que els arxius XML dels diaris ja no tenen a primera plana, però que les pàgines web conserven. De totes maneres, tot i tenir aquest historial, quan un usuari accedeixi a les notícies externes mitjançant el format de dades RSS no llegirà la informació d'aquest historial, sinó que accedirà directament, online, a les pàgines web d'origen, ja que si s'ha produït una novetat d'última hora podrà ser visualitzada. Per fer cerques directes la consulta serà online, en canvi per fer cerques avançades es farà les consultes a l'historial que hi haurà emmagatzemat a la base de dades.

2.3.6 Modificar una notícia

Per a modificar una notícia, només es poden modificar les notícies pròpies, el procediment és similar a l'anterior. Tan sigui per crear com per modificar una notícia, és necessari que es tracti d'un usuari amb permisos d'administrador. En el formulari ens apareix la informació desitjada, però ara es pot corregir. Podem canviar el títol, el subtítol, el cos de la notícia i el link, ja que la data i l'hora són camps que s'afegeixen de forma automàtica.

2.4 Esquema general

La vista general de la pàgina principal tindrà un esquema segons es mostra a la Fig. 5, dividint la pantalla en 3 frames horitzontals. Un primer estàtic format per

dues columnes, un segon dinàmic format també per dues columnes molt més extens i un final estàtic. A continuació es fa la descripció detallada de cada zona de la pàgina.

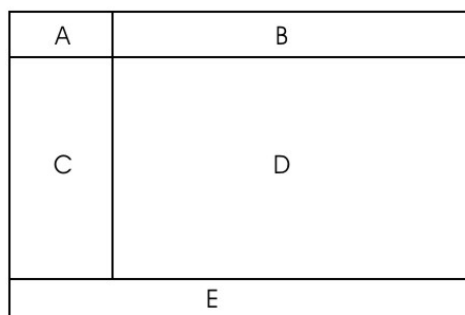


Fig. 5 – Esquema de rss_inici_noticies.php

2.4.1 Frame estàtic superior

A: S'obtenim el nom de l'usuari i es mostra per pantalla.

B: Hi ha els diferents temes que es poden tenir activats (esports, política, economia, societat i cultura). Els que siguin accessibles apareixeran actius i els altres inactius, ressaltant el que hi hagi en el moment.

2.4.2 Frame dinàmic central

C: Hi haurà un mòdul de cerca d'informació. Quan s'introdueix el text que es vol buscar i es prem a Ok busca la paraula pels títols i els textos dels diaris que hi ha actius i quan troba la paraula o paraules guarda la notícia a la base de dades. Un cop finalitzada la cerca, mostra a la zona D les notícies que s'han trobat. A sota del mòdul de cerca, apareixeran els diaris que es té seleccionats.

D: Aquesta és la zona on interactuarem amb més freqüència amb l'aplicació. Aquí és on es mostren les notícies, tant les dels diaris generals, com les notícies de les cerques que s'hagin fet. També és on apareixerà la informació quan es vulgui fer modificacions de les seleccions, això ho detallarem més endavant.

2.4.3 Frame estàtic inferior

E: La última zona de la pantalla és una barra de menús inferior, on es troba l'accés a modificar la selecció dels diaris i el botó per sortir i desconnectar-se de l'aplicació. Un cop l'usuari s'hagi desconnectat el codi generat amb anterioritat s'elimina de la base de dades.

3. ARQUITECTURA

El projecte que tenim entre mans es basa en una aplicació web, per tant, per una banda hi haurà un navegador que permetrà visualitzar la informació i modificar la que vulguem, i per una altra un servidor d'internet amb un servidor d'aplicacions, on es guardarà la informació en una base de dades i es processaran les operacions amb un llenguatge de programació. Tot seguit es detalla l'arquitectura i els seus components.

3.1 Plataforma WAMP

Primer de tot es descriu el que s'anomena arquitectura WAMP, ja que és la que s'utilitzarà per la nostra aplicació i la que apareix exemplificada amb els logotips corresponents a la Fig. 6, les sigles del qual fan referència a Windows (sistema operatiu) [3], Apache (servidor web) [4], MySQL (base de dades) [5, 6] i PHP (llenguatge de programació) [7]. Es pot afirmar que es tracta d'una plataforma per a la construcció de llocs web utilitzant programari lliure, tot i que per ser lliure del tot s'hauria d'haver utilitzat el sistema operatiu Linux passant-se a denominar arquitectura LAMP. També existeix la plataforma MAMP, aquest cas el sistema operatiu funciona amb Mac.



Fig. 6 – Plataforma WAMP

3.1.1 Windows (Sistema operatiu)

La majoria d'ordinadors que hi ha al mercat funcionen amb el sistema operatiu de Microsoft i per desenvolupar aquesta aplicació ens trobem en aquest cas, conté la plataforma Windows XP, per tant s'ha hagut d'instal·lar el paquet

WampServer 2.0 [8], el qual conté les versions següents de les eines que formen part de l'arquitectura WAMP: Apache 2.2.6, MySQL 5.0.45 i PHP 5.2.5. També conté el gestor phpMyAdmin, descrita més endavant.

3.1.2 Apache (Servidor Web)

Apache HTTP Server és un servidor HTTP (de pàgines web) de codi obert multiplataforma desenvolupat per Apache Software Foundation. Quan va començar el seu desenvolupament l'any 1995, es basava inicialment en el codi del popular NCSA HTTPd 1.3 (Centre Nacional d'Activitats de Supercomputació, Universitat de Ullinois), però més tard es reescriuria completament.

El seu nom prové de que, originalment, consistia bàsicament en un conjunt de pedaços a aplicar al servidor NCSA: en anglès, *a patchy server* (un servidor apedaçat). Apache presenta entre d'altres característiques missatges d'error altament configurables, bases de dades d'autenticació i negociació de continguts, però va ser criticat per la manca d'una interfície gràfica que ajudi a configurar-lo.

Des d'abril de 1996 Apache ha estat el servidor HTTP més popular a la World Wide Web; tot i això, ha experimentat una declinació de la seva quota de mercat, perduda en major part contra Microsoft Internet Information Services i .NET.

L'arquitectura utilitzada és client/servidor, és a dir, el client fa una sol·licitud o petició a l'equip servidor i aquest l'atén. El client fa d'interfície amb l'usuari i es comunica amb el servidor web.

Unes funcionalitats remarcables [9] d'apache són la utilització de mòduls, la creació de hosts virtuals o els mecanismes d'autenticació dels usuaris. Tot seguit se'n descriuen les seves funcionalitats i característiques principals.

La utilització de mòduls, és una de les característiques importants d'Apache, ja que no totes les instal·lacions de servidors web necessiten les mateixes funcionalitats. Existeixen dos tipus de mòduls: els que es compilen de forma estàtica, cada cop que es compila Apache2 i els que es carreguen dinàmicament, això permet fer canvis sense haver de recompilar tot el programa de nou.

Els arxius `.htaccess` permeten als usuaris que no tenen permisos modificar la configuració i així poder exercir algun control sobre el comportament de la seva part del servidor. Cada cop que se sol·licita un document s'executa `.htaccess`. De totes maneres no són molt recomanables ja que això farà alentir el funcionament del sistema, perquè cada cop farà la comprovació a tots els directoris i a més pot afectar a la seguretat, ja que es permeten modificacions a usuaris que no tenen permisos assignats.

Treballar amb hosts virtuals consisteix en executar més d'un lloc web en el mateix servidor. Té com a avantatges la versatilitat de crear diferents llocs web configurables; el preu, ja que només es necessita una sola màquina; una configuració del sistema serveix per tots els servidors web. Apache soporta dos tipus de hosts virtuals, basats en noms i en IP.

Pel que fa a l'autenticació d'usuaris existeixen dos mètodes: Bàsic o simple, i Digest. En el primer l'usuari introdueix el seu login i password al navegador i s'envien al servidor sense xifrar. En el segon cas s'envien el login i el password xifrats, però no tots els navegadors web el suporten. Són mètodes que controlen l'accés als recursos però no protegeixen la informació intercanviada. El control d'accés es restringeix amb les directives `Allow` i `Deny` que permeten o deneguen l'accés al servidor.

Totes les proves que s'han fet amb l'aplicació han sigut sobre el servidor d'HTTP (HyperText Transfer Protocol), Apache 2.2.6 el qual permet treballar

amb el llenguatge de programació PHP, cosa que no permeten per exemple els servidors web amb plataforma Windows.

3.1.3 MySQL (Base de Dades)

Al servidor també hi ha allotjada una base de dades. Tenim instal·lat el gestor MySQL 5.0.45, ja que és un dels sistemes gestors de bases de dades més populars desenvolupats sota la filosofia de codi obert i lliure i ha esdevingut molt popular gràcies a la seva velocitat en executar consultes. La base de dades guardarà per una banda tota la informació de les persones que es registren al portal (nom, login, password, i els diaris que es volen tenir seleccionats) i per una altra banda totes les notícies que es van guardant per poder mostrar amb posterioritat.

MySQL està escrit en C i C++, provat en un ampli rang de compiladors diferents. Funciona en diverses plataformes, Unix, Linux, Windows, etc... Les funcions SQL estan implementades usant una llibreria altament optimitzada i han de ser tan ràpides com sigui possible. Normalment no hi ha reserva de memòria després de tota la inicialització per a consultes. A continuació es fa una breu descripció d'algunes característiques de MySQL

És molt popular, el fet que molta gent la utilitzi i que estigui present en gran part de les webs, fa que els proveïdors de software intentin que les seves aplicacions siguin compatibles amb aquesta base de dades, podent trobar moltes aplicacions per a MySQL.

És simple, és fàcil d'aprendre comparada amb altres base de dades, no és necessari dedicar-li molt temps o diners en formar treballadors o en contractar a gent qualificada. Al ser senzilla, també és ràpida.

És barata, ja que MySQL pertany a open source, codi lliure, com s'ha explicat a l'inici, això fa que a l'inici del projecte el gasto vagi dedicat a altres parts que poden necessitar-ho, en comptes de gastar diners amb llicències.

Té suport, un dels inconvenients que solen presentar els clients cap al codi lliure és que no disposa de suport, en canvi MySQL ofereix llicència amb serveis de suport i manteniment, a més també és possible obtenir atenció tècnica 24/7.

És flexible i escalable, disposa de molts tipus d'emmagatzematge, podent escollir el que més convingui i els procediments guardats permeten escalabilitat.

Finalment també té llibreries natives per a llenguatges, ofereix llibreries per llenguatges com PHP i Ruby.

Hi ha altres gestors de base de dades, com podria ser Oracle, que tenen una robustesa i maduresa segurament superior a MySQL, però té el gran inconvenient que per a poder usar tot el potencial es necessita una llicència, que econòmicament no es pot assumir, i alhora les versions de prova que apareixen al mercat no permeten desenvolupar l'aplicació amb garanties d'èxit.

3.1.4 PHP (Llenguatge de Programació)

Del llenguatge de programació PHP s'utilitza la versió 5.2.5 ja que hi ha algunes funcions i classes que no són compatibles amb versions anteriors. Per exemple hi ha la funció *file_get_contents* que s'utilitza per llegir els arxius XML i necessita una versió superior a la 4.3.0 de PHP, però també hi ha les funcions *preg_match_all*, o *array_slice* que com a mínim necessiten la 4. El PHP és un llenguatge de programació de servidor independent del navegador, serveix per comunicar-lo amb una altra aplicació, el qual el fa poderós ja que utilitza els recursos propis del servidor. Es tracta d'un llenguatge de programació

interpretat, lliure i força popular, utilitzat per a generar contingut dinàmic al web.

En comptes de moltes pàgines HTML, les pàgines de PHP contenen HTML amb codi arrelat que fa "alguna cosa". El codi de PHP s'adjunta a instruccions de processament de començament i final especials `<?php` i `?>` que li permet saltar a dins i fora del mode PHP. Tot seguit en tenim un exemple a la Fig. 7.

```
<?php
$var1 = "2";
$var2 = $var1 + $var1;
$text = "Hola món";
echo "$var1 + $var1 = $var2<br>";
echo "$text<br>";
?>
```

Fig. 7 – Exemple de PHP

El que distingeix PHP de per exemple JavaScript és que aquest últim està al costat del client, en canvi amb PHP el codi s'executa al servidor, i l'HTML que genera és llavors enviat al client. El client rebria els resultats de funcionar aquella escriptura, però no sabria quin era el codi subjacent. Pot fins i tot configurar el seu servidor de webs per processar tots els seus arxius de HTML amb PHP, i llavors no hi ha realment cap manera que els usuaris puguin saber el codi que genera aquestes pàgines, tan sols veuran el codi de les pàgines un cop creades per mostrar-les als clients. Alguns avantatges de l'ús de la programació amb PHP són que és extremadament simple per a un nouvingut, però a la vegada ofereix molts trets avançats per a un programador expert o professional.

Tot seguit es mostra el procés de funcionament de les pàgines que funcionen amb PHP. La Fig. 8 mostra el procés que hi ha des que l'usuari demana la informació fins que la rep per pantalla.

Primer de tot l'usuari a través de l'aplicació que funciona mitjançant un navegador demana la informació que vol rebre, per exemple, obrir les notícies del diari Avui, pitjant sobre el logo del diari desitjat. Aquesta petició es fa al servidor d'internet. Aquest servidor utilitza el llenguatge PHP per fer les consultes necessàries a la base de dades i generar una pàgina html que serà la que s'enviarà a l'usuari. Aquest procés de PHP pot ser, i de fet serà, diferent segons què demani l'usuari. Si es treballa en el cas d'obrir les notícies del diari Avui, primer de tot es fa una consulta a la base de dades per obtenir l'adreça del diari d'on s'extrauran les notícies. Un cop es té l'adreça es farà una consulta via RSS a la web per tal de rebre la informació. En aquest cas les notícies es van escrivint una sota l'altra formant una pàgina HTML. D'aquesta manera es van obtenint les pàgines de forma dinàmica, cada cop que l'usuari obre un diari diferent no hi ha pàgines diferents al servidor, sinó que és el mateix arxiu *rss_inici_noticies.php* que genera diferent informació segons quin diari se li assigni.



Fig. 8 – Exemple del funcionament amb PHP

3.2 Software utilitzat

Un cop s'ha mostrat l'arquitectura principal a continuació es detallarà el software utilitzat per poder realitzar i desenvolupar l'aplicació.

3.2.1 phpMyAdmin

Per crear l'estructura de la base de dades s'ha utilitzat el gestor phpMyAdmin, la versió 2.10.1, ja que és una eina escrita amb php per manejar l'administració de MySQL a través de pàgines web, utilitzant un navegador. D'aquesta manera s'aconsegueix una optimització pel que fa a temps dedicat al tractament de la base de dades, sigui a la creació com a la modificació de registres i taules.

3.2.2 Macromedia Dreamweaver

Per poder desenvolupar el llenguatge de programació PHP s'ha utilitzat el Macromedia Dreamweaver 8 [10]. Aquest software és una potent eina de desenvolupament de pàgines web, creada inicialment per editar pàgines web però que mica en mica ha anat ampliant la seva oferta, incorporant la programació de JavaScript o CSS i també suporta llenguatges de programació de servidor, com ara l'esmentat PHP que nosaltres utilitzem. Aquesta eina facilita la visualització del codi, marcant amb colors diferents conceptes bàsics de la programació: amb negre les variables, amb blau les funcions i les operacions, amb verd els while, for, if, amb granat el que està entre cometes, amb vermell els nombres i l'obertura i tancament de PHP, per posar uns quants exemples. La programació queda ben estructurada ja que es veu contínuament l'arbre d'arxius i directoris que formen l'aplicació i el codi de la pàgina que s'està treballant, a més no afegeix codi brossa a les instruccions. Permet manipular la informació de la pàgina des de la vista de Disseny o es pot crear codi complex des de la vista de Codi. A més està connectat amb el navegador que es tingui per defecte i permet visualitzar les proves directament. És molt robust.

3.2.3 Corel Photo Paint

Hi ha diversos programes de tractament d'imatges, segurament els més coneguts i alhora més potents del mercat són el Paint Shop Pro de Corel i el Photoshop d'Adobe. En aquest cas l'estàndard en les empreses acostuma a ser el Photoshop, tot i així per portar a terme l'aplicació s'ha utilitzat el paquet

Corel Draw Graphics Suite 11 [11], i més exactament el Corel Photo-Paint ja que les funcionalitats necessàries les compleix perfectament. Aquest paquet gràfic ofereix gran flexibilitat i una àmplia varietat d'eines de dibuix, i suporta l'ús de capes, d'objectes, de filtres, efectes gràfics especials i drag-and-drop. És una potent eina de retoc fotogràfic. Tot el disseny visual de l'aplicació, ja siguin icones, botons dels menús, les imatges dels diaris, els títols i tota la interfície gràfica s'ha dut a terme amb el Corel Photo Paint 11, ja que és molt complet i permet fer tot allò necessari per tal de tenir una bona presència a la web. També s'ha fet amb aquesta eina el diagrama entitat-relació de la base de dades que apareix més endavant, al capítol 4 d'aquesta memòria.

3.2.4 Internet Explorer i Mozilla Firefox

Finalment el software principal que es necessita perquè l'aplicació pugui funcionar és el navegador, el browser. Per tal que sigui el màxim de compatible possible s'han fet proves amb el Mozilla Firefox 2, i amb l'Internet Explorer 7. Principalment s'ha treballat amb el Mozilla Firefox ja que és el que està enllaçat amb el Dreamweaver, i per tant el que s'obre quan es fan les proves de programació. De totes maneres amb posterioritat també s'ha obert l'aplicació amb l'Internet Explorer per veure si les funcionalitats produeixen algun error o tenen alguna incompatibilitat.

3.3 Format de dades RSS

RSS és una família de formats de canals web XML utilitzat per a publicar continguts actualitzats freqüentment com ara llocs de notícies, weblogs o podcasts, i per mitjà del qual es pot compartir la informació i usar-la en altres llocs web o programes. És en essència una sindicació de continguts.

L'acrònim s'usa per als següents sistemes de comunicació estàndard:

- Rich Site Summary (RSS 0.91)
- RDF Site Summary (RSS 0.9 i 1.0)

- Really Simple Syndication (RSS 2.0)

Els programes que llegeixen i presenten fonts RSS de diferents procedències es denominen agregadors. Aquests es poden trobar de dues maneres diferents, per mitjà d'una subscripció en un lloc web, i des d'un software dissenyat per llegir aquests continguts.

Quan s'obté la informació des del navegador no es necessita cap software addicional. En canvi també es pot distribuir contingut sense necessitat d'un navegador, utilitzant un software específic dissenyat per llegir aquests continguts.

Gràcies als agregadors o lectors de canals (programes o llocs que permeten llegir les fonts RSS) es poden obtenir resums de tots els weblogs que es desitgi des de l'escriptori del sistema operatiu, programes de correu electrònic o per mitjà d'aplicacions web que funcionen com a agregadors. No és necessari obrir el navegador i visitar tantes webs.

Però RSS no només serveix a l'usuari per rebre informació que altres ofereixen, sinó que també és d'utilitat per mostrar continguts nous de la pròpia web a altres internautes. Per això es necessita que l'usuari creï el seu propi feed i l'actualitzi freqüentment amb notícies novedoses sobre el tema que hagi escollit.

El procés de sindicació de continguts es converteix d'aquesta manera en un circuit de doble sentit en el que tots accedeixen i creen informació amb major facilitat i de forma més ràpida.

Els arxius XML tenen unes característiques concretes en quan a la seva estructura. Normalment tots comencen fent referència a la versió que tenen i a quina codificació es basa les seves notícies.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<?xml version="1.0" encoding="iso-8859-1"?>
```

Seguidament apareixen les etiquetes amb informacions diverses, sempre estan compostes per parelles, una primera que obre i una darrera que tanca, amb el mateix nom però amb la barra a davant:

<channel></channel>: Principi i final del canal RSS

<item></item>: Bloc que engloba una sola notícia

<title></title>: Títol de la notícia

<link></link>: Enllaç a la notícia sencera que es troba al servidor de cada web

<description></description>: Breu descripció o resum del cos de la notícia

<pubDate></pubDate>: Data en que s'ha incorporat la notícia a la web

<category></category>: Categoria de la notícia

Dins d'aquestes etiquetes es troba el text que s'haurà d'agafar per tal de poder mostrar les notícies de forma correcte. Durant el període de cerca de diaris i configuració de cada arxiu s'han trobat dificultats que s'han hagut d'anar superant fent programacions específiques per cada un. Molts textos tenen sentències html d'obertura de negreta o cursiva però no tenen el tancament. Això fa que les especificacions afectin a les notícies següents si no s'arregla.

3.4 Interacció global

A continuació es presenta, la Fig. 9, un quadre amb l'esquema global de funcionament de l'arquitectura que s'utilitza per dur a terme aquesta aplicació.

A l'esquema es veu que l'usuari interactua amb l'aplicació mitjançant un navegador web, des d'aquest navegador s'accedeix al servidor d'HTTP, on hi ha l'aplicació en sí, l'interpret de PHP i la Base de Dades. Un cop s'ha fet una petició d'HTTP l'interpret de PHP llegirà un arxiu i farà les corresponents crides a la base de dades i/o a internet per mitjà del format de dades RSS. Finalment es retorna a l'usuari un document HTML que visualitzarà altre cop des del navegador.

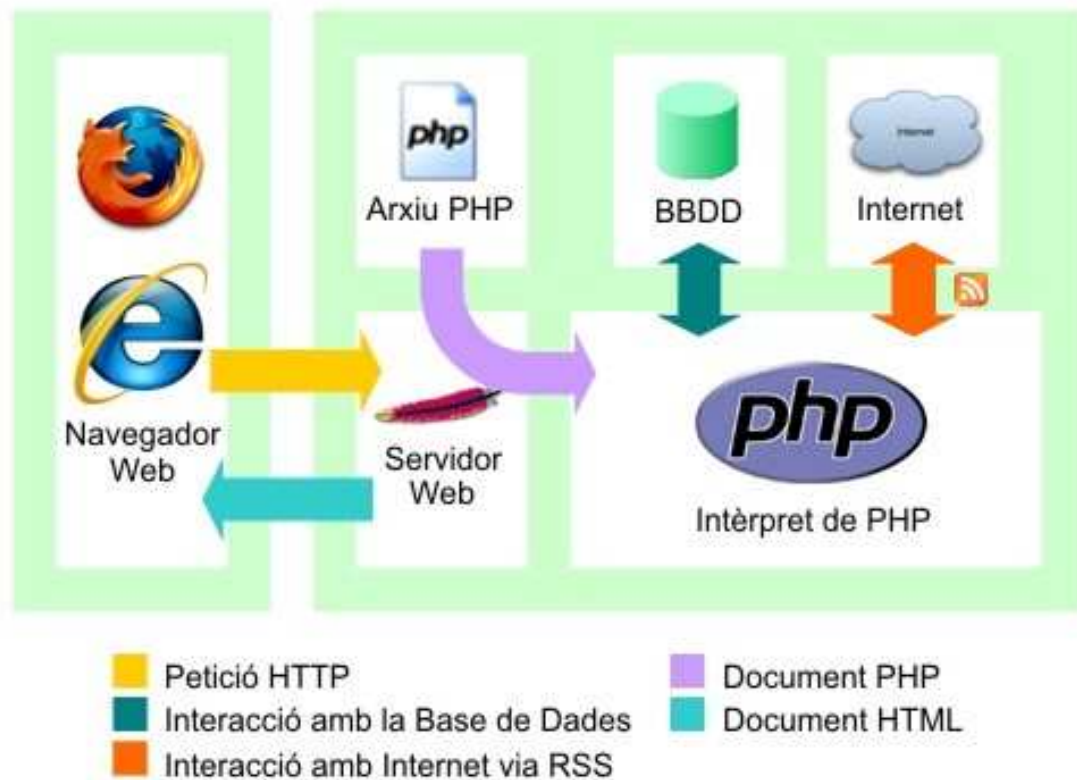


Fig. 9 – Interacció global

Tot seguit es mostra un exemple de funcionament.

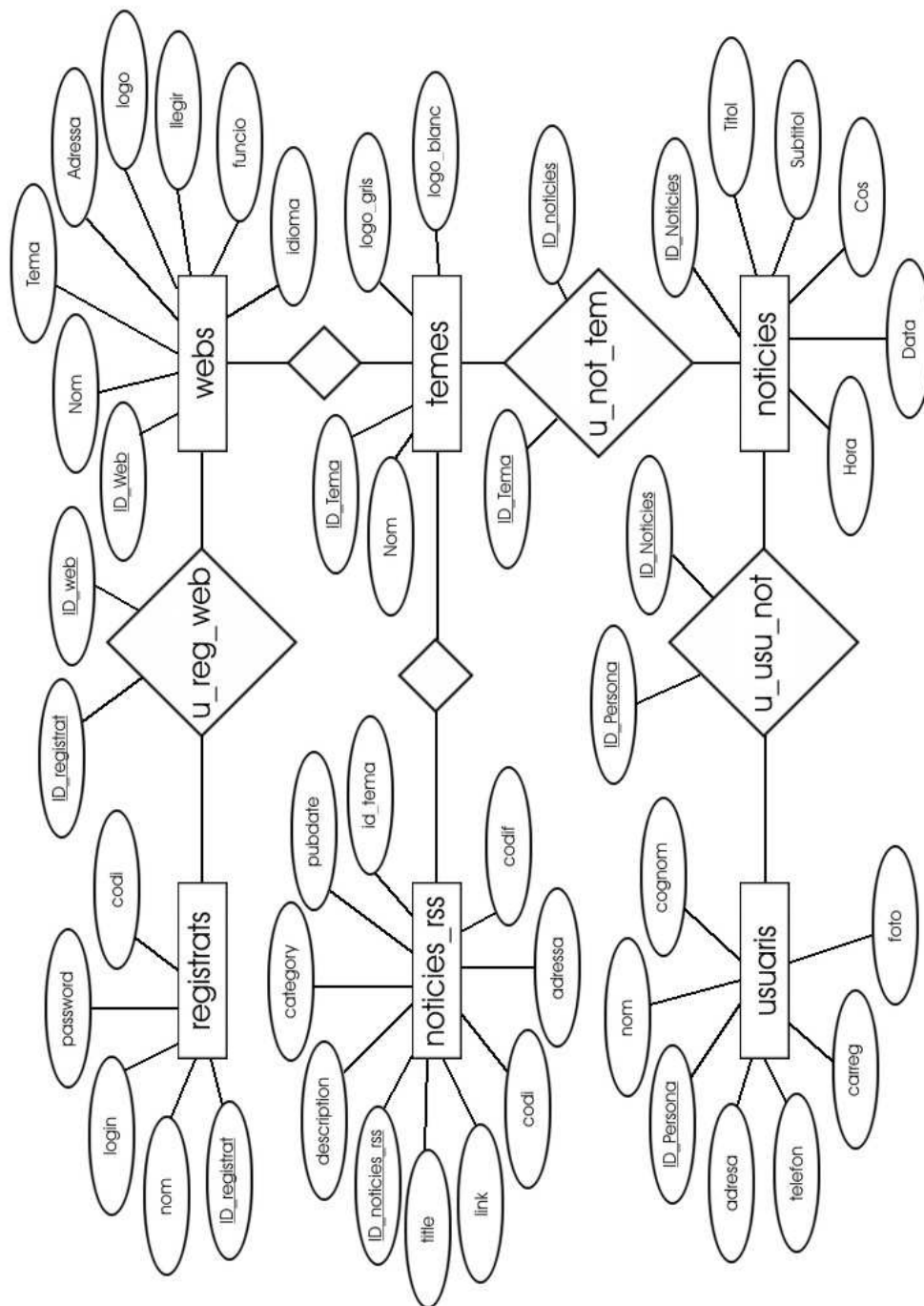
Un usuari accedeix a l'aplicació i introdueix el login i password des del seu navegador d'internet. Aquests són enviats via HTTP al servidor web, el qual amb l'interpret de PHP accedirà a la base de dades per comprovar si són correctes. Un cop validats, torna a accedir a la base de dades per llegir quins diaris té seleccionats l'usuari. Una vegada es tenen identificats els diaris que vol rebre l'usuari es fa un accés a Internet mitjançant el format de dades RSS, per crear amb posterioritat un document HTML i mostrar les notícies per pantalla.

Com es pot veure, la interacció entre l'interpret de PHP i la base de dades o internet és constant i repetitiva, ja sigui per fer comprovacions de logins i passwords o per fer consultes dels diaris.

4. DISSENY I IMPLEMENTACIÓ

4.1 Diagrama E-R

A continuació es fa una descripció del diagrama entitat-relació de la base de dades i tot seguit s'expliquen els arxius que formen part de cada funcionalitat.



4.1.1 Part d'usuari visitant

Les entitats que s'utilitzen quan entra un usuari visitant, que no està registrat, tan sols són *notícies* i *temes*, amb els seus respectius atributs. Els arxius que s'utilitzen aquí es troben al directori notícies i són els següents:

- inici_noticies.php (pantalla inicial amb un iframe central que crida a:)
- cerca_noticies.php (mostren les notícies de la cerca)
- opcions_avansades_normals.php (diverses opcions de selecció)
- lectura_noticies.php (mostren les notícies segons el tema)

4.1.2 Part d'usuari registrat

Un usuari registrat pot arribar a fer servir totes les entitats de la base de dades, excepte la d'*usuaris* que és exclusiva dels administradors, tot i que amb més freqüència utilitzarà *registrats*, *webs* i *notícies_rss* amb els *temes* corresponents. Sempre tenint permisos de lectura, mai d'escriptura. Els arxius que formen part d'aquesta estructura es troben al directori RSS i són els següents:

- rss_inici_noticies (menú general amb un iframe central que crida a:)
- rss_suport.php (quan es vol llegir un diari es crida al proper arxiu)
- rss_llegeix_nomdiari.php (accedeix a l'arxiu XML i imprimeix les notícies)
- opcions_avansades.php (diverses opcions de selecció)
- modifica.php (i modifica2.php, 3 i 4. Per canviar els diaris seleccionats)
- lectura_rss.php (s'hi accedeix per fer cerques de paraules)
- generar_codi.php (genera el codi de l'usuari)
- coloca_data.php (igual a la data de les notícies AA-MM-DD)

4.1.3 Part d'usuari administrador

Un usuari amb privilegi d'administrador serà l'únic que podrà utilitzar l'entitat *usuari*, que està mal anomenada, ja que hauria de dir-se administrador. Aquest és el que pot crear nous registres a *notícies* i alhora l'únic que pot crear nous registres a *webs*, totes elles lligades amb l'entitat *temes*. Els arxius es troben al directori admin:

- inici_admin.php (pàgina principal d'administració)

opcions.php (iframe central amb les opcions als arxius següents)
afegir_diaris.php (per afegir un nou diari d'accés RSS)
../noticies/insert_noticies.php (per inserir notícies pròpies)
../noticies/llicar_noticies.php (per seleccionar una notícia pròpia)
../noticies/modifica_noticia.php (per modificar-la)

4.2 Disseny funcional

4.2.1 Introducció login i password

L'aplicació comença demanant un login i un password per tal d'entrar amb una informació predeterminada que cadascú té guardada. Aquest login i password es comproven a la base de dades i si són correctes generen un codi de 25 dígit (números i lletres). Aquest codi serà el que s'utilitzarà durant la sessió. Si el login o el password introduïts no existeixen a la base de dades o són incorrectes mostra un missatge d'error. En cas que un usuari nou es vulgui donar d'alta ho farà pitjant al botó corresponent i omplint el formulari que apareix, amb el nom, login i password. En aquest cas s'afegirà a la taula *registrats* un nou registre amb totes les dades noves. El codi es genera cada cop que s'introdueix el login i el password, i s'elimina quan es tanca la sessió.

Quan s'entra a la pàgina principal, Fig. 11, tan sols es passa el codi d'usuari, i automàticament es busca un diari dels guardats a la base de dades, que quedarà ressaltat a la part esquerra de la pantalla, i el primer tema per defecte, en un altre color al frame superior. Tenint aquesta informació s'omple la pàgina amb les notícies corresponents. Cada cop que es crida la pàgina principal es fan una sèrie de comprovacions per veure què s'està modificant, segons si existeix valor a les variables \$tema i \$diari.

Si no es passa cap valor a \$diari ni a \$tema vol dir que s'està en el cas inicial, que s'accedeix a la pàgina principal un cop introduït el login i el password o que s'acaben de fer unes modificacions.



Fig. 11 – Captura de rss_inici_noticies.php

Si es té valor a \$diari però no a \$tema vol dir que s'està canviant de diari, que s'ha apretat al nom d'un altre diari a la part esquerra de la pantalla. Per tant es modificaran les notícies.

Si es té valor a \$diari i a \$tema significa que s'està canviant la temàtica però del mateix diari que es té seleccionat. En aquest cas també es modificaran les notícies a la zona central de la pàgina en correspondència amb allò indicat.

4.2.2 Visualització de notícies pròpies

Les notícies pròpies estan guardades a la base de dades. Són introduïdes per un usuari administrador i podran ser visualitzades per tothom qui ho desitgi. Aquestes notícies són les que podran veure els usuaris que no estiguin registrats a l'aplicació. Per tal de veure les notícies es farà una cerca a la base de dades amb els paràmetres que es desitgin. Per defecte apareixeran les últimes 5 notícies que hi hagi guardades al servidor com a notícies inicials. Posteriorment es poden escollir els diferents temes que apareixen al menú superior de la

pàgina on es visualitzaran les 5 darreres notícies de cada tema. Per tal de poder accedir a notícies anteriors es farà usant l'aplicació d'opcions avançades, on es podrà escollir entre unes dates concretes, un tema determinat, i/o paraules que es desitgin. Amb una consulta SQL obtindrem la informació necessària, ja que totes les notícies estan guardades a la base de dades.

4.2.3 Visualització de notícies externes

Tal i com s'ha descrit en el capítol anterior, el format de dades RSS és un estàndard per poder obtenir informació i fer-ne difusió pública. Tot seguit es mostra el procediment a seguir per obtenir la informació, les notícies externes, i col·locar-les a la nostra aplicació.

Existeix un arxiu .php per cada diari que s'afegeix a l'aplicació, ja que cada pàgina web codifica les notícies a la seva manera, tot i ser un llenguatge per poder ser utilitzat des de l'exterior, no són exactament iguals. El format del nom de l'arxiu és el següent: el text *rss_llegeix_* seguit del nom del diari i finalitzat amb l'extensió .php. Per tant l'arxiu per visualitzar les notícies de La Vanguardia és *rss_llegeix_lavanguardia.php*. Cada arxiu que l'administrador crea al nostre servidor té una funció amb un nom identificatiu del diari en qüestió, el mateix que apareix al nom de l'arxiu, així se la pot cridar diverses vegades i és més funcional i eficient. El format de la funció és el següent:

```
function lavanguardia($codi,$enll,$cerca=0,$paraula=null)
```

Consta d'un primer paràmetre amb el codi de l'usuari, un segon paràmetre amb l'arxiu XML que s'ha de llegir, un tercer paràmetre per saber si s'està fent una cerca d'una paraula o simplement es vol llegir la informació i finalment un darrer paràmetre que per defecte és nul i que en cas contrari assenyalarà quina paraula s'està buscant a l'aplicació. Aquesta cerca està explicada al proper apartat, Cerca per paraula.

Els arxius XML que es llegeixen contenen un títol, un enllaç i un text o descripció, segons quins arxius tenen una data, la majoria, i alguns tenen categoria. Per tant, per començar, s'han d'indicar els marcadors que s'utilitzen per buscar-los a l'arxiu i fer-ne uns patrons. D'aquesta manera s'agafarà la informació que es troba continguda dins d'aquests patrons. Cada arxiu té les seves particularitats:

Per llegir la web del Regió7

```
$patron_titulo='#\<title\>(.*?)\</title\>#is';
$patron_link='#\<link\>(.*?)\</link\>#is';
$patron_description='#\<description\>(.*?)\</description\>#is';
$patron_pubdate='#\<pubDate\>(.*?)GMT\</pubDate\>#is';
```

Per llegir la web de El Periódico

```
$patron_titulo='#\<title\>\<![CDATA\[([.*?])\]]\>\</title\>#is';
$patron_link='#\<link\>(.*?)\</link\>#is';
$patron_description='#\<description\>\<![CDATA\[([.*?])\]]\>\</description\>#is';
$patron_pubdate='#\<pubDate\>(.*?)\</pubDate\>#is';
```

Com es veu en la captura de patrons, sempre es troben uns tags, unes etiquetes que són obligatòries i que són les següents: <title>, <link> i <description> tot i que també es podria anomenar estàndard la següent perquè gairebé sempre apareix l'etiqueta <pubDate>. Segons quin diari s'estigui tractant conté a més una segona codificació entre l'obertura i el tancament de les etiquetes abans esmentades, i que normalment tenen la forma següent:

```
<title><![CDATA[Text que forma el títol]]></title>
```

A continuació es fa la cerca d'aquests patrons dins del text amb el codi que apareix tot seguit i es va guardant la informació en arrays. Segons la web que s'està obrint apareixen al principi de l'arxiu etiquetes de <title>, <link>, <description> o <pubDate> amb informació referent a la pàgina web d'origen que no aporta informació útil per la nostra aplicació. Si és aquest el cas es fa una

eliminació d'aquestes primeres posicions dels arrays, aconseguint tenir la primera notícia a la primera posició.

```
preg_match_all ($patron_titulo, $xml, $arr_titulos);
preg_match_all ($patron_link, $xml, $arr_links);
preg_match_all ($patron_description, $xml, $arr_descriptions);
preg_match_all ($patron_pubdate, $xml, $arr_pubdates);
```

```
$arr_titulos[1] = array_slice($arr_titulos[1],0);
$arr_links[1] = array_slice($arr_links[1],2);
$arr_descriptions[1] = array_slice($arr_descriptions[1],0);
$arr_pubdates[1] = array_slice($arr_pubdates[1],1);
```

Un cop s'han guardat totes les notícies en els diferents arrays s'han de mostrar per pantalla. Aquí apareix un problema de codificació de les dades. Cada arxiu XML amb contingut RSS té una codificació de les seves dades i aquesta pot ser de dos tipus:

```
charset=UTF-8
charset=iso-8859-1
```

La majoria d'arxius estan codificats amb iso-8859-1 però n'hi ha que tenen la codificació UTF-8 i s'han de tractar diferent, ja que si no es fes així no s'acceptarien cerques amb paraules accentuades o es guardaria la informació de forma errònia, ja que la nostra Base de Dades està codificada amb iso-8859-1.

Exemple del codi font al moment de mostrar les notícies de La Vanguardia:

```
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>La Vanguàrdia</title>
</head>
```


Quan es vol mostrar les notícies es fa un recorregut, amb un foreach, pels diferents arrays on estan guardades, i es van mostrant per pantalla el títol, el text, l'enllaç, la data i la categoria si en té, amb un tipus de lletra determinat, mida i color. Segons hi hagi una codificació o una altra s'han de fer diferents conversions, que es mostren tot seguit amb dos exemples.

Arxiu codificat amb iso-8859-1

```
printf("<font face='Verdana' size='1' color='#777777'>%s</font><br>", $arr_pubdates [1][$i],0);
printf("<font face='Verdana' size='2' color='#777777'><b><a href=%s
      class='notis'>%s</a></b></font><br>", $arr_links[1][$i],$v,0);
printf("<font face='Verdana' size='2' color='#000000'>%s</font><br><br>",
      $arr_descriptions[1][$i],0);
```

Arxiu codificat amb UTF-8

```
printf("<font face='Verdana' size='1' color='#777777'>%s</font><br>", $arr_pubdates[1][$i],0);
printf("<font face='Verdana' size='2' color='#777777'><b><a href=%s
      class='notis'>%s</a></b></font><br>", $arr_links[1][$i],
      htmlspecialchars_decode(stripslashes($v)),0);
printf("<font face='Verdana' size='2' color='#000000'>%s</font><br><br>",
      htmlspecialchars_decode(stripslashes($arr_descriptions[1][$i])),0);
```

4.2.4 Modificació de la selecció

S'accedeix a aquesta part de l'aplicació per dues vies diferents. La primera és si s'entra a l'aplicació per primera vegada, la qual cosa significa que un usuari s'acaba de donar d'alta, la segona opció és quan es vol modificar la selecció de diaris establerta.

Tant si s'accedeix per una via o per l'altra quan s'entra a l'aplicació, apareix la pantalla inicial amb la zona central amb un combobox amb 3 opcions per escollir els diaris segons tres opcions predefinides (Idioma, Diari i Tema). Aquestes opcions permeten seleccionar els diaris dels quals es rebrà informació, es visualitzaran les notícies. Internament hi ha programats una sèrie de

combobox's, també anomenats select's relacionats entre ells amb iframes que es van fent visibles a mesura que es va escollint allò desitjat.

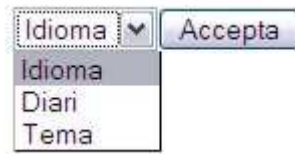


Fig. 11 – Primer select

El codi següent està inserit a l'arxiu modifica.php i es crida en el moment que es vol modificar una selecció o que s'entra a l'aplicació per primera vegada, genera el select de la Fig. 11, el primer d'una sèrie d'enllaçats entre si.

```
$codi = $_SESSION['codif'];
if ($_POST['accepta'])
{
    $opcions[0]="Idioma";
    $opcions[1]="Diari";
    $opcions[2]="Tema";
    printf("<form id='form1' name='form1' method='post' action='modifica.php'>");
    printf("<select name='select'>");
    for ($a=0; $a<3; $a++)
    {
        if ($opcions[$a] == $_POST['select'])
            printf("<option value='$opcions[$a]' selected='selected'>");
        else
            printf("<option value='$opcions[$a]'>");
        printf("%s</option>", $opcions[$a]);
    }
    printf("</select>");
    printf("<input type='submit' name='accepta' value='Accepta'>");
    printf("</form>");
    $link = Conectarse();
    $valor = $_POST['select'];
    printf("<iframe src='modifica2.php?valor=$valor' frameborder='0' align='middle' height='435' width='370'></iframe>", 0);
}
else
{
    printf("<form id='form1' name='form1' method='post' action='modifica.php'>");
    printf("<select name='select'>");
    printf("<option value='Idioma'>Idioma</option>");
    printf("<option value='Diari'>Diari</option>");
    printf("<option value='Tema'>Tema</option>");
    printf("</select>");
    printf("<input type='submit' name='accepta' value='Accepta'>");
    printf("</form>");
}
```

Si se selecciona la opció d'Idioma, apareixerà un segon combobox, tal com es mostra a la Fig. 12, amb els idiomes que hi ha als diaris, aquests idiomes no són

fixes sinó que es busquen a la base de dades depenent dels diaris que s'hagin introduït.



Fig. 12 – 2n select

Seleccionant l'idioma desitjat, apareixerà un tercer combobox, un exemple es mostra a la Fig. 13, amb els diaris que hi ha en la llengua seleccionada. Finalment seleccionant el diari apareixeran una sèrie de checkbox's on escollir les temàtiques que es vulgui del diari. Aquestes temàtiques poden arribar a ser fins a sis de diferents (esports, política, economia, societat, cultura i general), i en funció del diari seleccionat n'hi haurà uns o altres. Si un diari és exclusivament esportiu tant sols tindrà el checkbox d'esports, un diari d'informació general pot contenir diverses temàtiques, per exemple La Vanguardia té esports, política, economia, cultura i general, però no té societat. En canvi El Periódico conté tota la llista excepte cultura i general.

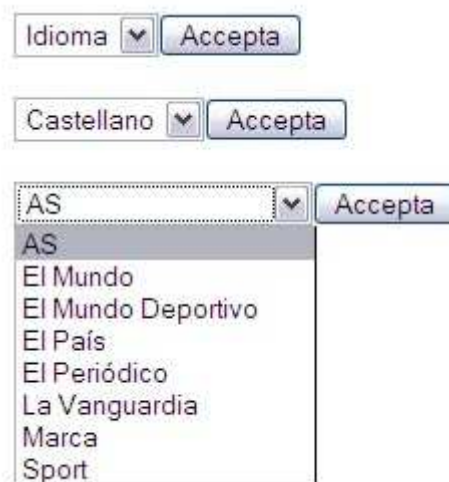


Fig. 13 – 3r select

Seleccionant Diari, apareixerà un segon combobox, mostrat a la Fig. 14, amb tots els diaris que hi ha a la base de dades i escollint el diari posteriorment apareixerà una llista de checkbox's amb els temes de cada diari.

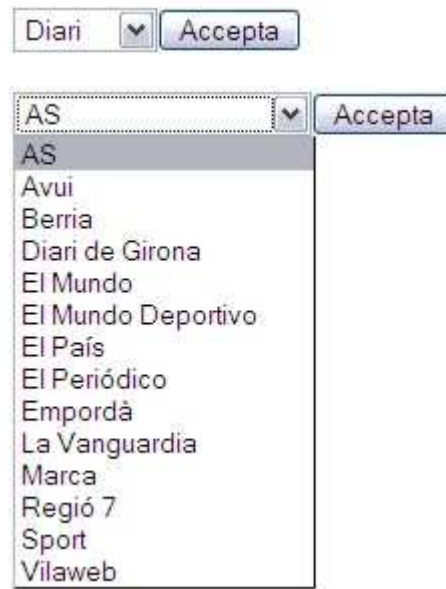


Fig. 14 – Diaris que tenim

Seleccionant Tema, tal com es mostra a la Fig. 15, apareixen les temàtiques que hi ha predefinides (General, Esports, Política, Economia, Societat i Cultura). Escollint una d'aquestes temàtiques s'obté un llistat de checkbox's amb tots els diaris que compleixen la condició.



Fig. 15 – Diaris segons el tema

Un cop s'han seleccionat els diaris que es volen tenir a la selecció particular de cadascú s'ha de pitjar el botó que apareix a la part inferior dels select's i que porta per nom "Modifica" el qual actualitza la base de dades amb la nova selecció, però sense tancar la pantalla. D'aquesta manera s'aniran fent les modificacions pertinents i modificant la base de dades alhora. Quan s'hagi

finalitzat amb les modificacions i es vulgui sortir del menú a la part esquerra de la pantalla hi ha un botó amb el títol “Finalitzar” que retorna a l’usuari a les notícies amb els canvis que s’hagin fet.

Fig. 16 – Modificació de la selecció

Per actualitzar la base de dades es fa la següent comprovació quan es pitja el botó Modifica: Si el checkbox està seleccionat es mira si aquesta pàgina està guardada a la base de dades, si ho està no es fa res, sinó s’afageix un registre nou a la taula u_reg_web amb l’identificador de l’usuari i l’identificador del diari corresponent. Igualment es fa la comprovació amb els checkbox’s que no estan seleccionats, es comprova si estan guardats, si és afirmatiu s’eliminen, en cas contrari es deixen fora.

Si s’accedeix a l’aplicació per segona vegada o posteriors, l’usuari ja no entrarà a la part de modificar les notícies sinó que es llegirà de la base de dades la informació que es té guardada i es busca la selecció d’uns determinats diaris i es visualitzaran directament les notícies. Un cop es tenen les notícies en pantalla

sempre es poden modificar pitjant al botó “Modificar Selecció” del menú inferior.

4.2.5 Cerca de paraules

Si el que s’està fent és una cerca d’una paraula, Fig. 17, pitjant a Ok s’accedeix a l’arxiu *rss_inici_noticies.php*, però introduint un valor de 1 a la variable \$cerca i el text desitjat a la variable \$paraula. Aquests valors faran que al frame central de la pàgina es faci una crida a l’arxiu *lectura_rss.php* amb el valor de les dues variables i el codi d’usuari.

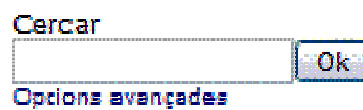


Fig. 17 – Mòdul de cerca

L’arxiu *lectura_rss.php* comença eliminant de la base de dades qualsevol notícia que es tingui guardada amb el codi de l’usuari i continua fent una consulta a la base de dades per tal de trobar tots els diaris que l’usuari té seleccionats, i poder fer-ne una passada amb posterioritat. D’aquesta manera s’accedeix als arxius *rss_llegeix_X.php*, sent la X el diari que es té seleccionat com si volguéssim mostrar les notícies per pantalla, però fent una comprovació per si la paraula existeix al text. Si és cert que l’ha trobat, es fa una inserció a la base de dades amb la notícia trobada i el codi de l’usuari. Tot seguit es mostra un tros del codi de la comprovació de la cerca, sent \$v el títol i \$text el cos de la notícia:

```
$enllas = addslashes($arr_links[1][$i]);
$text = addslashes($var_desc);
$pubdate = $arr_pubdates[1][$i];
if ($cerca == 1)
{
    if ((ereg($paraula, $v)) || (ereg($paraula, $text)))
    {
        $tem = mysql_query("select Tema from webs where Adressa = '$enllas', $link);
        while($row_tem = mysql_fetch_array($tem))
        {
            $val_tem = $row_tem["Tema"];
        }
    }
}
```

```

    }
    $tit = addslashes($v);
    mysql_query("insert into noticies_rss (title,link,description,pubdate,id_tema,codi,adresa,codif) values
('$tit','$enllas','$text','$pubdate','$val_tem','$codi','$enll','8859')", $link);
  }
}

```

El codi té petites modificacions segons el diari que s'obre, tal i com s'ha comentat anteriorment, ja que uns tenen el (codif) value (8859) o (UTF8). Els que tenen la codificació UTF8 no s'ha de fer la conversió del títol i per tant la línia de codi *\$tit = addslashes(\$v)*; no apareix.

Al finalitzar totes les cerques, des de l'arxiu *lectura_rss.php* es fa una segona consulta a la base de dades, aquest cop a la taula *noticies_rss* de totes aquelles notícies que tinguin adjunt el codi d'usuari correcte. Finalment es fa un volcat de les notícies guardades a la base de dades. Aquesta cerca es fa online, accedint als arxius XML dels portals web al moment de fer la cerca i copiant les notícies que continguin la paraula desitjada a la base de dades de forma temporal.


4.2.6 Cerca per opcions avançades


Cada dia es guarden totes les notícies de tots els diaris que hi ha a l'aplicació, independentment de les que tingui seleccionades cada usuari, s'explica amb més detall al capítol 4.2.8. D'aquesta manera l'aplicació tindrà un historial de notícies prou extens per poder fer-ne ús quan es vulgui i poder accedir a notícies passades que els arxius XML dels diaris ja no tenen, però els servidors web sí. Per tant aquestes cerques no seran als arxius XML sinó que serà una consulta interna a la base de dades pròpia les quals enllaçarà amb posterioritat a les notícies complertes.


Quan s'entra a l'aplicació com a usuari registrat es pot fer una cerca de notícies particularitzada. Internament es tracta d'un select a la base de dades amb una sèrie de condicionants depenent de les restriccions que se li afegeixin tal com es mostra a la captura de la Fig. 18 on apareixen els camps a escollir.


A les notícies pròpies també s’hi pot fer una cerca avançada, però en aquest cas no hi haurà el camp d’idioma. En aquest cas també es fa una consulta de la base de dades amb les notícies pròpies, a la taula notícies.

Seleccionar entre dates:

Inici: 

Final: 

Idioma: 

Tema: 

Paraula:

Fig.18 – Opcions avançades

4.2.7 Inserir nous diaris a l’aplicació

Per tal de poder afegir un nou diari a l’aplicació s’ha de tenir permís d’administrador. Els passos que s’hauran de seguir són els següents.

- Crear dues imatge amb el logotip del diari de 23 pixels d’alçada per màxim 170 píxels d’amplada, una a color i una segona amb escala de grisos.
- Crear un arxiu amb el nom “*rss_llegeix_Nom del diari.php*” amb les característiques i el format que s’han vist al capítol 4.2.3.
- Omplir el formulari que apareix a l’aplicació on es demana les dades del diari per introduir-les a la Base de Dades.

Aquest últim punt, el formulari de l’aplicació, consta de:

Nom del diari, es mostra un llistat amb tots els diaris que hi ha a la base de dades per si s’hi vol afegir un tema nou. També apareix un camp per introduir el nom d’un nou diari si no existeix.

Idioma del diari, aquí també apareix un llistat dels idiomes que hi ha a l'aplicació i un camp de text per si és un idioma nou.

Un camp per col·locar-hi la temàtica de la pàgina XML, en aquest cas el combobox està predefinit i només pot ser un dels sis temes següents: esports, política, economia, societat, cultura o general.

Adreça web on es troba l'arxiu XML des d'on s'agafarà la informació mitjançant RSS.

El path del logo en color i un altre amb el logo en escala de grisos.

Finalment es demana el path de l'arxiu creat en el segon punt del procés, el qual té la forma d'accés a l'arxiu XML, la codificació i les comprovacions pertinents.

Un cop s'han fet aquests tres passos ja apareixerà el nou diari a la selecció. Es pot anar a modificar la selecció dels diaris i automàticament s'afegirà a la llista, podent seleccionar-lo per visualitzar-ne les notícies.

4.2.8 Guardar notícies a l'aplicació

La intenció de bon principi era fer un accés a l'arxiu *actualitzar_noticies.php* de forma automàtica cada dia arribada una certa hora, però no ho he pogut desenvolupar. Per internet he trobat que des de Linux es pot fer utilitzant el crontab però des de Windows no he aconseguit fer funcionar el VisualCron. Per tant, es fa un accés a aquest arxiu de forma manual diàriament. El procés és molt senzill.

1. Es fa una consulta a la base de dades per trobar tots els diaris que hi hagi introduïts.
2. Es fa una passada per tots els diaris amb la variable \$modifica igual a 2, així que introduirà a la base de dades les notícies però sense codi.
3. Un cop estiguin totes les notícies guardades a la taula *noticies_rss* es fa una còpia a la taula temporal, idèntica a la *noticies_rss*, agrupant-les pel camp *link*, que ens servirà per eliminar les notícies repetides.
4. Es buida la taula *noticies_rss*.

5. Es copia les dades de la taula temporal a `noticies_rss`.
6. Finalment es buida la taula temporal.

A la taula `noticies_rss` s'hi guarda els següents camps:

`ID_noticies_rss`: Un identificador únic per distingir les notícies, es guarda de forma autonumèrica.

`Title`: El títol de la notícia.

`Link`: Enllaç a la web on es troba la notícia sencera.

`Description`: Text del cos de la notícia.

`Pubdate`: Data de publicació de la notícia. En aquest camp s'hi fa una transformació ja que la data arriba de diferents maneres segons sigui un diari o altre. A la base de dades es guarda en el següent format: AA-MM-DD (any-mes-dia) i es fa a través d'una funció guardada a l'arxiu *coloca_data.php*. Dels diaris que s'han tractat la informació ens arriba de quatre maneres diferents: "2008-05-26T22:00:01", "Fri, 30 May 2008 00:55:00", "Sun, 8 Jun 2008 10:10:00 +0200" i finalment el diari Avui no té el camp `pubdate`, en aquest cas s'agafa la data de l'ordinador i s'introdueix.

`Id_tema`: Identificador del tema de la notícia.

`Adressa`: Adreça web on es troba l'arxiu XML d'on es llegeix la notícia.

`Codif`: Codificació en que es guarda la notícia (UTF-8 o iso-8859-1)

5. CONCLUSIONS

Un cop s'ha finalitzat el projecte, l'aplicació web de notícies, es revisaran els objectius que s'havien proposat al principi, sortint-ne molt satisfets. Tots ells s'han complert. Anem a fer-ne un breu repàs:

- Distingir 3 zones al portal, una d'usuaris visitants, una d'usuaris registrats i una tercera d'administració.

És la primera pàgina que apareix a l'aplicació, amb els diferents camins a escollir, depenent dels permisos que tingui l'usuari que accedeix al portal. Tal i com mostren els propers tres punts, cada usuari té unes característiques ben definides i una informació particularitzada.

- La zona d'usuaris visitants només ha de permetre visualitzar les notícies que el nostre portal contingui a la nostra base de dades.
- La zona d'usuaris registrats ha de permetre importar notícies de diverses pàgines web mitjançant la tecnologia RSS.
- La zona d'usuaris amb perfil d'administradors ha de permetre introduir notícies segons la necessitat que tinguem.

Dins la zona d'usuaris registrats es pot seleccionar la informació que es rep segons els següents objectius fixats.

- Seleccionar les pàgines web d'on es volen treure les notícies.
- Seleccionar temàtiques específiques per rebre només unes certes notícies.

La tecnologia RSS és relativament nova i moderna, la qual cosa ha fet que a mesura que anava fent la programació de l'aplicació, les diferents pàgines web dels diaris, canviessin la codificació dels seus arxius XML.

Un aspecte negatiu que he trobat un cop feia la cerca d'informació i l'aplicació en sí, és la falta de fotografies en els documents XML. Tan sols el diari Sport, conté de tant en tant algun enllaç a youtube la qual cosa permet veure vídeos directament des de l'aplicació. En un altre diari, l'Avui, apareix una etiqueta `<im:image>` que sembla ser de fotografia, però curiosament l'enllaç sempre és a un mateix lloc, a una imatge irrellevant. Aquest fet m'ha decebut bastant perquè la inclusió d'imatges hauria donat més dinamisme a l'aplicació i visualment hauria sigut més atractiva.

Exceptuant aquesta decepció, estic força satisfet del resultat final perquè jo soc de les persones que obren diversos diaris per veure diferents opinions de les notícies i d'aquesta manera està tot centralitzat. Tot i així no totes les notícies que emeten els diaris apareixen als documents XML, sobretot quan separen la informació en temàtiques. Depenent de les classificacions de cada diari, pot ser que no aparegui a la nostra aplicació. De totes maneres crec que el resultat és força satisfactori, ja que a mesura que anava introduint diaris anaven apareixen dificultats diverses sobretot particularitats de cada un que s'havien de tractar de forma individual. Tot i així com acostuma a succeir, després de les dificultats el resultat és més satisfactori.

Una de les coses que m'han sorprès força és amb l'alegria que alguns diaris tracten les pàgines XML. La majoria estan ben estructurades, però alguns diaris deuen tenir automatitzat el procés d'execució de l'arxiu sense massa restriccions. Això fa que per exemple els diaris del Grupo Zeta (Sport o El Periódico) tinguin certes notícies amb etiquetes d'obertura de negreta, cursiva o centrat però que no tinguin les de tancament. Per tal que no afecti a les notícies següents s'ha hagut de fer un control de tags al títol i les descripcions de les notícies cada cop que s'hi accedeix, per tal d'afegir-los si en falta algun.

De la mateixa manera que passa amb les etiquetes s'ha hagut de fer un control sobre el diari La Vanguardia, ja que dins l'etiqueta `<description>` al final del

text hi ha sempre una imatge d'un píxel de costat que no porta a enlloc. Per tant s'ha hagut de fer un control i quan es troba el text < s'elimina de l'array.

L'aplicació s'ha provat amb els dos navegadors majoritaris al mercat, tal i com s'ha descrit al capítol 3 d'arquitectura, el Firefox i l'Explorer. Depenent de si l'aplicació sobre des d'un o altra navegador hi ha lleugeres diferències en quan a la presentació ja que l'explorer té tendència a aixamplar els espais i els marges, en canvi el Firefox ho fa tot més reduït, potser s'ha de fixar algun paràmetre que un dels dos navegadors agafa per defecte i l'altra no. Però el que més sobte és que el Firefox no mostra totes les imatges que hauria de mostrar, però ho fa de forma aleatòria i sense cap sentit. Al refrescar la pàgina potser apareixen totes, potser en falta una, després una altra, i així en diverses pantalles. És un problema que no he pogut descobrir la solució, ja que no segueix cap seqüència lògica ni falla sempre al mateix lloc i en les mateixes condicions.

S'ha fet un estudi sobre el creixement de les pàgines que es guarden a la base de dades i els resultats obtinguts durant 12 dies del mes de maig és que es reben una mitjana de 630 notícies diàries, per un total de 7550 notícies aproximadament. Tot seguit apareix un parell de gràfics, Fig. 19, amb el nombre total de pàgines de cada temàtica que hi ha actualment introduïts a l'aplicació i el percentatge que representa.

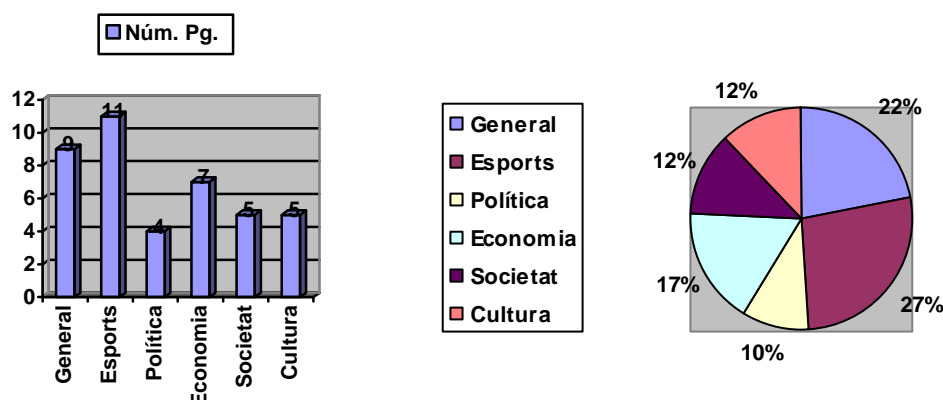


Fig. 19 – Nombre total i percentual de pàgines introduïdes per buscar.

A continuació apareixen dues gràfiques, Fig. 20, amb el resultat de les pàgines obtingudes de llegir els arxius XML. Com es pot veure les pàgines que més creixen són les d'esports, ja que en quant a pàgines d'aquesta temàtica representava el 27% del total i en canvi rep el 39% de les notícies que arriben. Això es deu a que tenim 4 diaris d'àmbit esportiu i que generen gran quantitat de notícies. Les notícies de societat són les que reben la davallada més forta. Segons si introduïm nous diaris i depenent de les temàtiques podríem anivellar els percentatges.

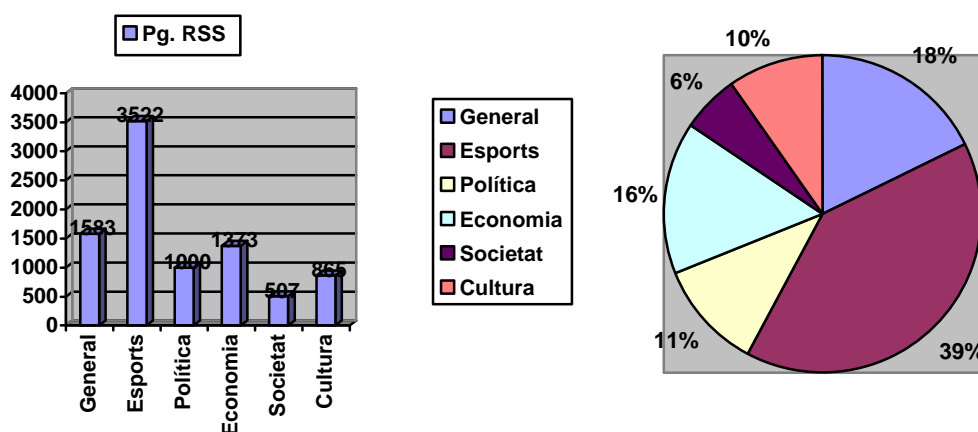


Fig. 20 – Nombre total i percentual de pàgines trobades via RSS.

La base de dades només creix per la taula de notícies_rss a un ordre d'aproximadament 7,5MB cada 10000 notícies.

5.1 Treball futur

Segurament en un futur no molt llunyà es podran passar via RSS fotografies i/o vídeos, per tant s'hauria de fer una modificació del codi de lectura dels arxius per adaptar-se a les noves necessitats.

Es podria fer unes aplicacions adaptades a les necessitats de l'usuari final, ja que cada cop hi ha més classificacions en quan a arxius XML, un diari esportiu

pot tenir classificacions d'esports, d'equips, de categories, de múltiples disciplines. Un diari d'informació general, pot arribar a tenir RSS de pobles, territoris, temàtiques, etc... Per tant, es podrien fer aplicacions a mida de l'usuari.

En quan a la tecnologia es podria aplicar una mica més de funcionalitat amb Javascript, en comptes dels iframes concatenats, això facilitaria l'execució a l'usuari ja que els canvis podrien ser instantanis i sense necessitat d'acceptar cada cop a un botó.

Estèticament podria guanyar atractiu si s'incorporessin les notícies d'última hora en moviment en una sola línia entre el menú superior i el frame central. S'hauria de fer un refresc constant a tots els diaris, la qual cosa alentiria el funcionament general de l'aplicació, però es podria buscar alguna solució per optimitzar-ho i podria guanyar en eficiència.

Un aspecte que no he tingut en compte i me n'he adonat un cop finalitzada l'aplicació és la construcció d'un arxiu XML per les notícies pròpies. Tal i com s'ha dit en el capítol 3.3 els arxius RSS són de doble via, de lectura i escriptura, poder agafar-los i poder enviar-los. No ho he tingut en compte perquè primer vaig començar l'aplicació per la part de les notícies pròpies i no va ser fins que vaig implementar les notícies externes quan em vaig introduir amb el codi per recuperar informació dels arxius XML amb el format de dades RSS. Per tant una bona millora seria implementar aquesta funcionalitat per tal de poder compartir les notícies amb enllaços des d'altres pàgines, weblogs, etc...

BIBLIOGRAFIA

Bibliografia general

Cercador Google: <http://www.google.cat>

Gran Enciclopèdia Catalana: <http://www.enciclopedia.cat>

Bibliografia específica

[1] <http://ec.grec.net/lexicx.jsp?GECART=0094724>

Diccionari de la Gran Enciclopèdia Catalana

[2] <http://ca.wikipedia.org/wiki/RSS>

[3] <http://www.microsoft.com/spain/windows/default.msp>

Sistema Operatiu Windows XP

[4] <http://www.apache.org>

Servidor d'HTTP Apache

[5] PC Cuadernos técnicos Vol.4 - PHP y MySQL páginas web dinámicas

[6] <http://dev.mysql.com/doc/refman/5.0/es/index.html>

Base de dades MySQL – Manual

[7] <http://www.php.net>

Llenguatge de programació PHP

[8] <http://www.wampserver.com>

Paquet que instal·la el WampServer

[9] <http://observatorio.cnice.mec.es/modules.php?op=modload&name=News&file=article&sid=580&mode=thread&order=0&thold=0&POSTNUKESID=ca0859281619a596d7dd074f7d1eba21>

Apache 2.2: Servidor Web

[10] <http://www.adobe.com/es/products/dreamweaver/features/>

DreamWeaver8

[11] <http://www.corel.com/servlet/Satellite/us/en/Product/1150981051301>

Corel Draw Graphics Suite